Genet Sel Evol (1990) 22, 57-64

Original article

Molecular cloning and partial sequence characterization of the duplicated amylase genes from *Drosophila erecta*

L Bally-Cuif, V Payant, S Abukashawa, BF Benkel, DA Hickey*

Department of Biology University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

(Received 16 May 1989; accepted 15 November 1989)

Summary – Previous studies have shown that the α-amylase gene is duplicated in each of the 8 species comprising the *Drosophila melanogaster* species subgroup. In this study, a genomic library in phage lambda was prepared from the DNA of 1 species member of this subgroup, *D erecta*. This gene library was screened with an amylase-specific probe from the closely-related species *D melanogaster*. One of the cross-hybridizing phage clones was isolated, insert fragments were sub-cloned into plasmid vectors, and the identity of the cloned locus was verified by DNA sequencing. The results confirm that this species contains 2 very similar copies of the amylase-coding sequence, as does *D melanogaster*. Within the amylase-coding region, there is a 3% sequence divergence between the 2 species; this value rises to 6% if we consider only silent sites within the coding region. Sequence divergence between *D erecta* and *D melanogaster* in the non-coding intergenic region is approximately 15%.

These results constitute the first molecular sequence data available for protein-coding genes of *D erecta*. The data set allows us to calculate an estimated divergence time of more than 10 million years between the 2 species. This is consistent with the results of previous phylogenetic studies.

***Drosophila erecta* / α-amylase / molecular evolution / gene conversion**

Résumé – Clonage et caractérisation d'une séquence partielle des gènes dupliqués de l'amylase chez *Drosophila erecta* – *Des études antérieures ont montré que le gène α-amylase était dupliqué chez chacune des 8 espèces du sous-groupe* Drosophila melanogaster. *Dans la présente étude, une banque génomique a été préparée dans le phage lambda à partir de l'une des espèces membres du sous-groupe;* D erecta. *Cette banque a été criblée avec une sonde, provenant de l'espèce voisine* D melanogaster, *et contenant les séquences codant pour l'amylase. Un des clones hybridant à la sonde a été isolé, et ses fragments de restriction sous-clonés dans des plasmides. La nature des gènes a été vérifiée par séquençage. Les résultats obtenus confirment que* D erecta *contient, comme* D melanogaster, *2 copies très semblables de la région codante du gène α-amylase. Dans cette région, la divergence de séquence entre les 2 espèces est de 3%, et atteint 6% si l'on ne considère que les positions silencieuses des codons. Dans la région intergénique (non codante), la divergence est d'environ 15%. Ces résultats constituent les premières données de séquence pour des gènes codants pour une protéine chez* D erecta. *Ils nous permettent*

* Correspondence and reprints

d'estimer un temps de divergence de plus de 10 millions d'années entre les 2 espèces, ce qui est compatible avec les résultats des études phylogénétiques antérieures.

Drosophila erecta / α-amylase / évolution moléculaire / conversion génique

INTRODUCTION

Drosophila α-amylase genes have already been extensively studied (Levy *et al*, 1985; Gemmil *et al*, 1986; Doane *et al*, 1987; Langley *et al*, 1988; Hickey *et al*, 1989b), and they have proved to be particular interest, mainly because of their numerous encoded allozymes (Singh *et al*, 1982; Dainou *et al*, 1987), their complex pattern of regulation, including glucose repressibility (Benkel and Hickey, 1987), and their duplicated structure. Within natural populations of *D melanogaster*, the gene is duplicated (Levy *et al*, 1985; Gemmill *et al*, 1986), with the 2 copies, each 1.4 kb in length, being separated by approximately 4.8 kb of sequence. They are present in opposite orientation (Levy *et al*, 1985; Benkel *et al*, 1987) and are divergently transcribed (Boer and Hickey, 1986). A simplified map of this locus is presented in Fig 1.

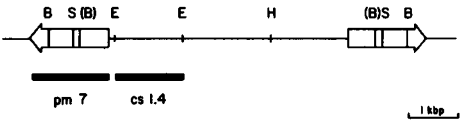

Fig 1. Molecular organisation of the α-amylase locus in *D melanogaster*. The coding regions of the gene are represented by boxes, and the arrows indicate the direction of transcription. According to the nomenclature of Gemmill *et al* (1986), the proximal gene is located on the left and the distal on the right. The restriction sites common to *D melanogaster* and *D erecta* are shown without parentheses (B: *Bam* H1; E: *Eco* R1; H: *Hind* III; S: *Sal* 1). The *Bam* H1 sites, absent from both gene copies in *D erecta*, are shown in parentheses. The probes used for the construction of the clones are indicated underneath as filled boxes.

The number and pattern of electrophoretic variants within each species and in interspecific hybrids (Dainou *et al*, 1987), coupled with the recent analyses of restriction maps (Payant *et al*, 1988), have shown that the duplication is not restricted to *D melanogaster*, but is also present in the 7 other species of the *melanogaster* subgroup. The α-amylase gene thus appears as a small multigene family, containing only 2 members. Because of the presence of this duplication in all the species of the subgroup, it can be considered as an evolutionarily ancient trait.

The 2 copies of the α-amylase gene from a strain of *D melanogaster* have already been partly sequenced (Boer and Hickey, 1986), and the 2 coding regions showed a high degree of sequence similarity. This, in addition to the close linkage of the 2 copies, raised the possibility of sequence recombinations between the 2 copies of the gene, even though their opposite orientation is considered as a more stable arrangement than tandem duplication. The construction of a restriction map for the α-amylase locus in each of the 8 species of the subgroup (Payant *et al*, 1988) proved that the general pattern of restriction sites was conserved between the species, but that a *Bam* H1 site located in the coding region was absent from the 2 copies of the

gene in both *D erecta* and *D teissieri* (Fig 1). If the mutation eliminating this *Bam* H1 site proved to be the same in the 2 copies of each species, this would support the hypothesis of recombination events occuring between the 2 copies of the gene within a locus.

We cloned both copies of the amylase gene from *D erecta* and sequenced part of the coding region (containing the *Bam* H1 location) for each gene copy. We then performed sequence comparisons both, within the α-amylase locus in the *D erecta* strain, and between the two species, *D erecta* and *D melanogaster*. This allowed us to assess sequence similarity between the 2 gene copies in *D erecta* and, also, to quantify the degree of sequence of divergence between *D erecta* and *D melanogaster*. These sequence data are discussed in relation to the results obtained in previous phylogenetic studies.

MATERIAL AND METHODS

The *D erecta* strain 220S, collected in 1980 in the Ivory Coast, West Africa, and provided by Dr J David, was used in the present study. Twelve flies of that strain were sampled and tested for their α-amylase allozymic pattern by polyacrylamide gel electrophoresis. Having verified the identity of the strain, genomic DNA was isolated from the descendants of these flies using a procedure described in detail elsewhere (Payant *et al*, 1988). After digestion of the DNA with *Sau* 3A, a genomic library was constructed in lambda phage EMBL 3, according to the manufacturer's instructions (Stratagene). Plaque lifts were prepared as described by Benton and Davis (1977), using nylon membranes (Biotrans; ICN Biomedicals). The membranes were hybridized to the pOR-M7 (coding region) *D melanogaster* amylase cDNA ^{32}p-labelled probe (Benkel *et al*, 1987); this probe is indicated by the bar labelled pm7 in Fig 1. The filters were then washed under conditions where only the recombinant phages, with at least 80% sequence similarity to the a amylase gene, would remain bound to the probe (for detailed procedure, see Payant *et al*, 1988). After elimination of the radioactivity, the same procedure was used to hybridize the membrane to the CS1.4a *D melanogaster* amylase ^{32}p-labelled probe (Hickey *et al*, 1988; and see Fig 1). Five clones hybridizing to both probes were identified. Their DNA was isolated and digested with *Sal* 1. A southern blot (Southern, 1975) of the digested DNAs was made, and probed with the pOR-M7 and CS.4a probes, as described above. A 6.4 kb *Sal* 1 fragment hybridized to the 2 probes. Based on its length and hybridization pattern, it was concluded that this fragment contained the 5-prime region of both copies of the α-amylase gene. This 6.4 kb fragment (see Fig 2) was then isolated, purified, and subcloned into the *Sal* site of the pUC-based vector, pIBI 21. The orientation of the subclones was determined by an *Eco* R1 digestion. Each subclone was then digested with *Sal* 1 and *Hind* III, and the *Sal* 1/*Hind* III fragments were further subcloned and inserted, in both orientations, into pIBI 24 and pIBI 25 plasmid vectors. The resulting subclones are presented in Fig 2 (inserts of 4 kb and 2.4 kb, respectively, for the proximal and the distal genes). These subclones were used for sequencing.

DNA sequences were obtained by the dideoxy chain termination technique (Sanger *et al*, 1977), using LKB Macrophor electrophoresis equipment. The se-

quences were assembled and analysed using a sonic digitizer and the Microgenie (Beckman) computer programs (Queen and Korn, 1984).

RESULTS AND DISCUSSION

For both amylase gene copies of *D erecta*, 500 bp of coding region were sequenced, along with a short portion (200 bp) of the non-coding intergenic region. These regions are indicated in Fig 2, and the sequencing results are presented in Fig 3.

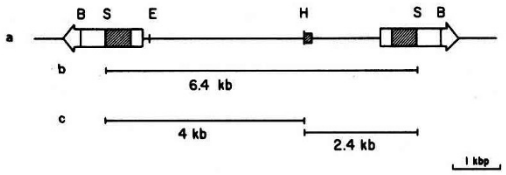

Fig 2. Localization of the *D erecta* subclones used in this study. (a) Restriction map of the α-amylase locus in *D erecta*. The symbols used are the same as those of Fig 1. The 500 bp sequenced are represented as hatched boxes within the coding region; the segment of the intergenic region which was sequenced is also indicated by a box, just to the right of the intergenic *Hin*d III site. (b) The 6.4 kb *Sal* 1 fragment inserted into pIBI 21 (in opposite orientations). (c) *Sal* 1/*Hin*d III restriction fragments inserted into pIBI 24 and pIBI 25, and used for sequencing.

Sequence similarity between the two D erecta *gene copies*

We found that the 2 gene copies in *D erecta* were 100% similar for the region studied. For instance, the mutation of the *Bam* H1 site in *D erecta* is due to an identical nucleotide substitution in the 2 gene copies. Several other mismatches separated *D melanogaster* and *D erecta* at the sequence level, but all of these polymorphisms were shared by both gene copies of *D erecta*. Thus, for that region, all the nucleotide substitutions that have occured since the *D erecta* and *D melanogaster* species divergence, have been incorporated into both gene copies of *D erecta*. Out of the 15 nucleotide substitutions that separated *D melanogaster* and *D erecta* in that region, 9 affected the 3rd base of a codon, and did result in any amino-acid substitution between the derived proteins. It is worth noting that even the silent substitutions were identical in both copies of the *D erecta* coding sequence. Thus, the near-identity of the two *D erecta* coding regions cannot be explained solely as result of strong selection acting in favour of the occurence of certain new amino acids in the amylase of *D erecta* (compared to that of *D melanogaster*), and thus, favouring the incorporation of the corresponding new nucleotides into both copies of the gene.

There are several possible explanations for the conservation of sequence similarity between the members of a gene family. In this case, neither very recent duplication, nor retrotransposition are plausible explanations, since (i), as was mentioned before, the duplication has been shown to be very ancient (Dainou *et al*, 1987; Payant *et al*, 1988), and (ii) a comparison of the upstream regions of the 2 genes copies in *D melanogaster* (Boer and Hickey, 1986) showed that the sequence identity exdented further than the coding regions. Thus, the 2 α-amylase gene copies within these species are most likely evolving in a concerted fashion (Hickey *et al*, 1990).

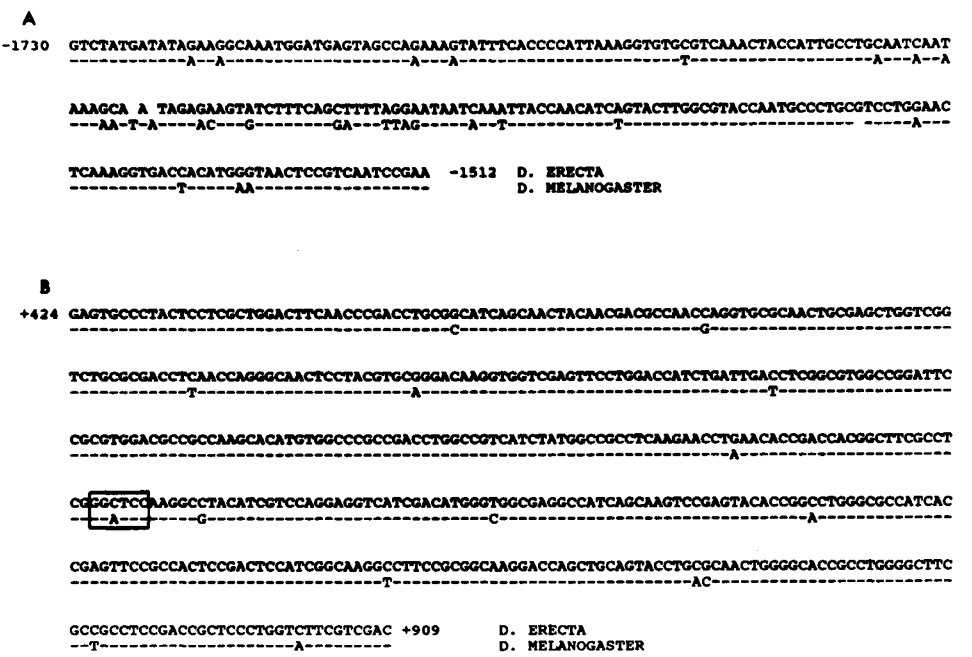

Fig 3. Comparison of the α-amylase coding and intergenic regions of *D erecta* with the homologous sequences of the *D melanogaster* strain, Oregon R (Boer and Hickey, 1986). (A) Sequence from the intergenic region (see Fig 2 for location of this region). (B) Sequence from the coding region (see Fig 2). Only the nucleotides that differ from *D erecta* are shown for *D melanogaster*. The mutated *Bam* H1 site (Payant *et al*, 1988; and see discussion in text) is boxed. Sequences are numbered relative to the translation initiation site. Although sequence data were obtained for both copies of *D erecta*, only a single coding sequence is shown for this species; this is because both gene copies had identical sequences for this portion of the coding region (see discussion in text).

Sequence divergence between D erecta *and* D melanogaster

An alignment of the *D erecta* amylase sequences with the proximal amylase of the *D melanogaster* strain Oregon R is presented in Fig 3. The time of divergence between the 2 species, *D erecta* and *D melanogaster*, can be estimated from these data. The short segment from the intergenic region (see Fig 2 and Fig 3A) shows approximately 15% divergence between the 2 species. This sequence is not subject to the constraints of natural selection acting on its coding capacity; therefore, it serves as a useful qualitative indicator of the considerable amount of divergence between the 2 species at the molecular level. Although other members of the species subgroup have been subjected to molecular analysis (*eg*, Bodmer and Ashburner, 1984), no previous molecular studies have been made on *D erecta*.

As one might expect, the coding region showed considerably less interspecies divergence than did the non-coding intergenic region. The 500 bp of coding sequence presented here (see Fig 3B), shows 3.3% nucleotide divergence between the 2 species; the length of those sequences, compared to that of the gene (1500 pp), is sufficient

to give a statistically valid approximation of the percentage of nucleotide divergence through the whole coding region.

Many analyses have already been carried out to estimate the divergence time between the 8 species of the *melanogaster* subgroup. Mainly based on allozymes and the 2-dimensional gel electrophoresis (Cariou, 1987), mitochondrial DNA (Solignac *et al*, 1986), ribosomal DNA and histone gene family organization (Coen *et al*, 1982), DNA sequence data (Bodmer and Ashburner, 1984) and polytene chromosomes banding patterns (see Lemeunier *et al*, 1986), the *D melanogaster* subgroup has been split into 3 complexes: the *D erecta – D orena* complex, the *D yakuba – D Teissieri* complex, and the *D melanogaster* complex (including *D melanogaster*, *D simulans*, *D mauritiana* and *D sechellia*) (see Lachaise *et al*, 1988; Lemeunier *et al*, 1986, for a review). Moreover, mainly because of its members' amylase allozyme patterns, mitochondrial DNA patterns and its inability to interbreed with species from the other complexes, the *D erecta – D orena* complex appears to be farthest apart from the 2 others, and the divergence time has been estimated to be as high as 15 million years.

Since some of our data concern the coding region of the gene, they can be used to confirm the above estimates of species divergence times. As expected, the level of divergence found between *D erecta* and *D melanogaster* (3%) is significantly higher than what was found between the α-amylase genes of different strains of *D melanogaster* (approximately 1%), indicating a greater divergence between the 2 species than within the latter (Boer and Hickey, 1986; Langley *et al*, 1988; and our unpublished data). The percentage divergence between the 2 species at silent sites (6.1%) can be used to estimate time since the divergence of *D erecta* and *D melanogaster*. If we use the rate of 5.6×10^{-9} changes per site, per year, proposed by Hayashida and Miyata (1983) for silent sites within exons, this percentage gives a divergence time of approximately 11 million years; the order of which is consistent with the previous estimates of divergence time. It should be noted, however, that this may be an underestimate of the real divergence time. For instance, as noted above, gene conversion can eliminate all substitutions, including those at silent size, between gene copies within a species, and this would result in a gross underestimates of the age of the duplication if we relied on sequence divergence at silent sites. Although gene conversion cannot, of course, occur between sequences which are separated in different lineages, other factors, such as biased GC content (see Hickey *et al*, 1989b), could reduce the observed rate of divergence at silent sites; this would result in a lowered estimate of the divergence time between species.

ACKNOWLEDGMENTS

This work was supported by an Operating Grant from NSERC Canada to DA Hickey and a Postdoctoral Fellowship from CNRS, France, to V Payant. We are grateful to Dr J David for providing stocks of *D erecta*.

REFERENCES

Benkel BF, Abukashawa S, Boer PH, Hickey DA, (1987) Molecular cloning of DNA complementary to *Drosophila melanogaster* α-amylase mRNA. *Genome* 29, 510-515

Benkel BF, Hickey DA (1987) A *Drosophila* gene is subject to glucose repression. *Proc Natl Acad Sci USA* 84, 1337-1339

Benton WD, David RW (1977) Screening lambda gt recombinant clones by hybridization to single plaques *in situ*. *Science* 190, 180-182

Bodmer M, Ashburner M (1984) Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. *Nature* 309, 425-430

Boer PH, Hickey DA (1986) The α-amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucl Acids Res* 14, 8399-8411

Cariou ML (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D sechellia* and *D orena*. *Genet Res* 50, 181-186

Coen E, Strachan T, Dover G (1982) Dynamics of concerted evolution species subgroup of *Drosophila*. *J Mol Biol* 158, 17-35

Dainou O, Cariou ML, David JR, Hickey DA (1987) Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. *Heredity* 59, 245-251

Doane WW, Gemmill RM, Schwartz PE, Hawley SA, Norman RA (1987) Structural organization of the α-amylase gene locus in *Drosophila melanogaster* and *Drosophila miranda*. *In: Isozymes: Current Topics in Biological and Medical Research*. Alan R Liss, New York, Vol 14, 229-266

Gemmill RM, Shwartz PE, Doane WW (1986) Structural organization of the Amy locus in seven strains of *Drosophila melanogaster*. *Nucl Acids Res* 14, 5337-5352

Hayashida H, Miyata T (1983) Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc Natl Acad Sci USA* 80, 2671-2675

Hickey DA, Benkel BF, Abukashawa S, Haus S (1988) DNA rearrangement causes multiple changes in gene expression at the amylase locus in *Drosophila melanogaster*. *Biochem Genet* 26, 757-768

Hickey DA, Bally-Cuif L, Abukashawa S, Benkel BF (1990) Concerted evolution of duplicated protein-coding genes in *Drosophila*. Submitted for publication

Hickey DA, Benkel BF, Magoulas C (1989b) Molecular biology of enzyme adaptations in higher eukaryotes. *Genome* 31, 272-283

Lachaise D, Cariou ML, Davis JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the *D melanogaster* species subgroup. *Evol Biol* 22, 159-225

Langley CH, Shrimpton AE, Yamazaki T, Miyashita N, Matsuo Y, Aquadro CF (1988) Naturally occurring variations in the restriction map of the Amy region of *D melanogaster*. *Genetics* 113, 619-629

Lemeunier F, David JR, Tsacas L, Ashburner M (1986) The *melanogaster* species subgroup. *In: The Genetics and Biology of Drosophila* (Ashburner M, Thompson JR, Carson HL, eds) Academic Press, London, Vol 3, 147-256

Levy JN, Gemmill RM, Doane WW (1985) Molecular cloning of α-amylase genes from *D melanogaster*. II. Clone organization and verification. *Genetics* 110, 314-324

Payant V, Abukashawa S, Sasseville M, Benkel BF, Hickey DA, David J (1988) Evolutionary conservation of the chromosomal configuration and regulation of amylase genes among eight species of the *Drosophila melanogaster* species subgroup. *Mol Biol Evol* 5, 560-567

Queen C, Korn LJ (1984) A comprehensive sequence analysis program for the IBM personal computer. *Nucl Acids Res* 12, 581-599

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74, 5463-5467

Singh RS, Hickey DA, David J (1982) Genetic differentiation between geographically distant populations of *D Melanogaster*. *Genetics* 101, 235-256

Solignac M, Monnerot M, Mounolou JC (1986) Mitochondrial DNA evolution in the melanogaster species subgroup of *Drosophila*. *J Mol Evol* 29, 31-40

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98, 503-517